

# *Correspondence Search Mitigation Using Feature Space Anti-Aliasing*

M. Veth, *Air Force Institute of Technology*  
M. Pachter, *Air Force Institute of Technology*

## **BIOGRAPHY**

Major Mike Veth is an Assistant Professor in the Department of Electrical and Computer Engineering at the Air Force Institute of Technology. His current research focus is on the fusion of optical and inertial systems. He received his Ph.D. in Electrical Engineering from the Air Force Institute of Technology, and a B.S. in Electrical Engineering from Purdue University. In addition, Major Veth is a graduate of the Air Force Test Pilot School.

Meir Pachter received the B.S. and M.S. degrees in aerospace engineering in 1967 and 1969, respectively, and the Ph.D. degree in applied mathematics in 1975, all from the Israel Institute of Technology.

He is a Professor of Electrical Engineering at the Air Force Institute of Technology. He has held research and teaching positions at the Israel Institute of Technology, the Council for Scientific and Industrial Research in South Africa, Virginia Polytechnic Institute, Harvard University and Integrated Systems, Inc. His current areas of interest include statistical signal processing, adaptive optics, inertial navigation, and GPS navigation.

Dr. Pachter received the Air Force Air Vehicle Directorate Foulis award in 1994 for his work on adaptive and reconfigurable flight control.

## **ABSTRACT**

Image-aided navigation techniques can determine the navigation solution (position, velocity, and attitude) by observing a sequence of images from an optical sensor over time. This operation is based on tracking the location of stationary objects in multiple images, which requires solving the correspondence problem. This is an active area of research and many algorithms exist which attempt to solve this problem by identifying a unique feature in one image and then searching subsequent images for a feature match. The correspondence problem is plagued by feature ambiguity, temporal feature changes, and also occlusions, which are difficult for a computer to address. Constraining the corre-

spondence search to a subset of the image plane has the dual advantage of increasing robustness by limiting false matches and improving search speed. A number of ad-hoc methods to constrain the correspondence search have been proposed in the literature.

In this paper, the correspondence problem itself is carefully analyzed from fundamental optical principles. This development results in a general temporal sampling constraint and also reveals the essential connection between the deleterious effects of temporal aliasing and the ambiguities which plague the correspondence search problem. This temporal image sampling constraint is expressed as a function of the navigation trajectory for elementary camera motions. The predicted sampling rates are on the order of those needed for adaptive optics control systems and require very large bandwidths. The temporal image sampling constraint is then re-evaluated by incorporating inertial measurements. The incorporation of inertial measurements is shown to reduce the required temporal sampling rate to practical levels, which evidences the fundamental synergy between image and inertial sensors for navigation and serves as the basis for a real-time, adaptive, anti-aliasing strategy.

## **INTRODUCTION**

It is well-known that optical measurements provide excellent navigation information, when interpreted properly. Optical navigation is not new. Pilotage is the oldest and most natively familiar form of navigation to humans and other animals. Mechanical instruments such as astrolabes, sextants, and driftmeters [17] have been used to make precision observations of the sky and ground to improve navigation performance for centuries.

The difficulty in using optical measurements for autonomous navigation, that is, without human intervention, has always been in the interpretation of the image, a difficulty shared with Automatic Target Recognition (ATR). Indeed, when celestial observations are used, the ATR problem in this structured environment is tractable and auto-

Report Documentation Page		Form Approved OMB No. 0704-0188
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.		
1. REPORT DATE <b>2007</b>	2. REPORT TYPE	3. DATES COVERED <b>00-00-2007 to 00-00-2007</b>
4. TITLE AND SUBTITLE <b>Correspondence Search Mitigation Using Feature Space Anti-Aliasing</b>		5a. CONTRACT NUMBER
		5b. GRANT NUMBER
		5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER	
	5e. TASK NUMBER	
	5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) <b>Air Force Institute of Technology, Department of Electrical and Computer Engineering, Wright Patterson AFB, OH, 45433</b>		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR'S ACRONYM(S)
		11. SPONSOR/MONITOR'S REPORT NUMBER(S)
12. DISTRIBUTION/AVAILABILITY STATEMENT <b>Approved for public release; distribution unlimited</b>		
13. SUPPLEMENTARY NOTES		
14. ABSTRACT <b>Image-aided navigation techniques can determine the navigation solution (position, velocity, and attitude) by observing a sequence of images from an optical sensor over time. This operation is based on tracking the location of stationary objects in multiple images, which requires solving the correspondence problem. This is an active area of research and many algorithms exist which attempt to solve this problem by identifying a unique feature in one image and then searching subsequent images for a feature match. The correspondence problem is plagued by feature ambiguity, temporal feature changes, and also occlusions, which are difficult for a computer to address. Constraining the correspondence search to a subset of the image plane has the dual advantage of increasing robustness by limiting false matches and improving search speed. A number of ad-hoc methods to constrain the correspondence search have been proposed in the literature. In this paper, the correspondence problem itself is carefully analyzed from fundamental optical principles. This development results in a general temporal sampling constraint and also reveals the essential connection between the deleterious effects of temporal aliasing and the ambiguities which plague the correspondence search problem. This temporal image sampling constraint is expressed as a function of the navigation trajectory for elementary camera motions. The predicted sampling rates are on the order of those needed for adaptive optics control systems and require very large bandwidths. The temporal image sampling constraint is then re-evaluated by incorporating inertial measurements. The incorporation of inertial measurements is shown to reduce the required temporal sampling rate to practical levels, which evidences the fundamental synergy between image and inertial sensors for navigation and serves as the basis for a real-time, adaptive, antialiasing strategy.</b>		
15. SUBJECT TERMS		

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT <b>Same as Report (SAR)</b>	18. NUMBER OF PAGES <b>12</b>	19a. NAME OF RESPONSIBLE PERSON
a. REPORT <b>unclassified</b>	b. ABSTRACT <b>unclassified</b>	c. THIS PAGE <b>unclassified</b>			

matic star trackers are widely used in astro-inertial navigation systems for long-range aircraft, space navigation, and ICBM guidance. When ground images are to be used, the difficulties associated with image interpretation are paramount. At the same time, the problems associated with the use of optical measurements for navigation are somewhat easier than ATR. Moreover, recent developments in feature tracking algorithms, miniaturization, and reduction in cost of inertial sensors and optical imagers, aided by the continuing improvement in microprocessor technology, motivates us to consider using inertial measurements to aid the task of feature tracking in image sequences and realize a tightly-coupled image-aided INS.

The methods are typically classified as either feature-based or optic flow-based, depending on how the image correspondence problem is addressed. Feature-based methods determine correspondence for “landmarks” in the scene over multiple frames, while optic flow-based methods typically determine correspondence for a whole portion of the image between frames using correlation techniques. A good reference on image correspondence is [13]. Optic flow methods have been proposed in the literature generally for elementary motion detection, in a somewhat structured environment focusing on determining relative velocity or angular rates for obstacle avoidance [7].

Feature tracking-based navigation methods have been proposed both for fixed-mount imaging sensors or gimbal mounted detectors which “stare” at the target of interest, similar to the gimballed infrared detector on some heat-seeking missiles. Many feature tracking-based navigation methods exploit knowledge (either *a priori*, through binocular stereopsis, or by exploiting terrain homography) of the target location and solve the inverse trajectory projection problem [1, 14]. If no *a priori* knowledge of the scene is provided, egomotion estimation is completely correlated with estimating the scene. This is referred as the structure from motion (SFM) problem. A theoretical development of the geometry of fixed-target tracking, with no *a priori* knowledge is provided in [16]. An online (Extended Kalman Filter-based) method for calculating a trajectory by tracking features at an unknown location on the Earth’s surface, provided the topography is known, is given in [5]. Finally, navigation-grade inertial sensors and terrain images collected on a T-38 “Talon” are processed and the potential benefits of optical-aided inertial sensors are experimentally shown in [18].

Many methods for solving the correspondence problem have been proposed in the computer vision literature. A popular algorithm is the Lucas-Kanade feature tracker [12], which relies on the premise of the invariance of the intensity field between images. It uses a template correlation algorithm to minimize the sum of squared differences (SSD) between image intensities. The algorithm typically

assumes a linear ( $x - y$  plane) motion model, but can be extended to optimize over affine or bilinear transformations. Other feature correspondence algorithms have been proposed which are invariant to rotations, scaling or both. (e.g., [10]) More robust feature tracking algorithms are typically computationally expensive and a designer must trade tracking robustness and accuracy for real-time performance.

## Current Correspondence Constraint Approaches

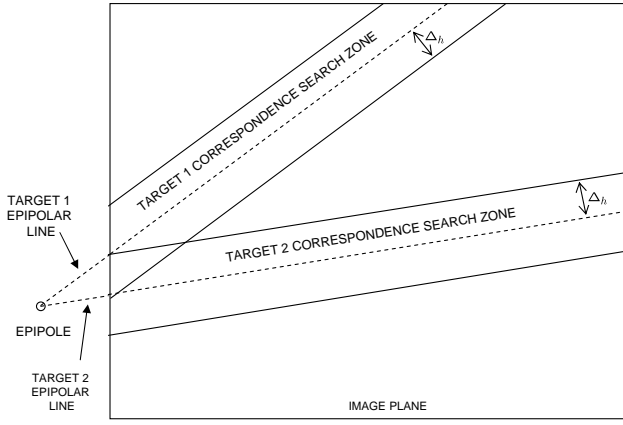
Exploiting inertial measurements to constrain the correspondence search has been proposed in the literature. In this section, two methods which exploit inertial measurements are discussed.

Bhanu and Roberts [3] utilize inertial measurements to compensate for rotation between images and to predict the focus of expansion in the second image. Once the second image is derotated and the focus of expansion is established, the correspondence between interest points is calculated using goodness-of-fit metrics. One relevant metric is the correspondence search constraint placed on each point. This constraint ensures each interest point lies in a cone-shaped region, with apex at the focus of expansion, bisected by the line joining the focus of expansion and the interest point in the camera frame at the first image time. While this constraint is not statistically rigorous, it does show the value of using inertial measurements to aid the correspondence problem.

Strelow also incorporates inertial measurements to constrain the correspondence search between image frames [20]. This constraint on the image search space is a similar concept to the field of expansion method proposed by Bhanu; however, Strelow generalizes the approach by exploiting epipolar geometry. The projection of an arbitrary point in an image is described by an epipolar line in a second image. All epipolar lines in an image converge at the projection of the focus of the complimentary image. Combining knowledge of the translation and rotation between images and the pixel location of a candidate target in the first image, a correspondence search can then be constrained to an area “near” the epipolar line. This approach is illustrated in Fig. 1.

Strelow’s method of using inertial measurements to constrain the correspondence search along an epipolar line is ad-hoc, since the search space is not defined statistically. This method could be improved by utilizing a stochastically rigorous development.

In previous publications, we have presented an approach which leverages the inertial measurements and any available terrain information to predict the locations and statistical uncertainty of features in a new image [24, 25]. Our



**Figure 1:** Correspondence search constraint using epipolar lines. Given a projection of an arbitrary point in an initial image, combined with knowledge of the translation and rotation to a second image, the correspondence search can be constrained to an area near the epipolar line. Note the epipole can be located outside of the image plane, as shown in this example.

goal in this article is to expand the stochastic constraint theory to an elemental level which is dependent on the inherent optical properties of the sensor. Analyzing the correspondence problem from this perspective reveals the parallel nature between feature correspondence searching and temporal sampling theory in signal processing which is well-understood. As a result, feature correspondence ambiguity is shown to be analogous to temporal aliasing. Thereby, sampling theory can be used to predict and mitigate/avoid the presence of aliasing in feature space.

In the next section, the theory of image sampling is developed from first principles, with particular attention to the anticipated issues with regard to temporal sampling.

## GENERAL IMAGE SAMPLING PROBLEM

The mathematical relationships governing spatial-temporal sampling are developed from basic optical and sampling theory. This development provides a theoretical basis which is used to develop temporal sampling constraints in subsequent sections.

### Image Sampling Considerations

A digital imaging device is, in essence, a sampler of light intensity patterns in three dimensions: two spatial and one temporal. Analyzing the effects of the sampling process on image sequences resulting from camera motion with due regard given to the motion's dynamics has very important implications on how to properly interpret image sequences to derive navigation information.

## Effects of Egomotion on Image Formation

As discussed in the previous section, the recorded image is a representation of the optical intensity patterns generated by a scene. The projection function is a function of the scene itself, the camera optical properties, and the pose (i.e., relative position and orientation) of the camera and scene. This strong coupling between camera pose and the image is the basis for the rapidly growing research efforts dedicated to exploiting images to determine changes in camera pose. In this section, the geometric projection function is developed using a pinhole camera model. This model will be used as a basis for quantifying the effects of egomotion and temporal sampling.

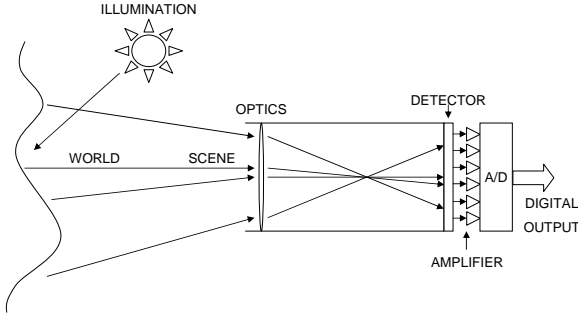
## Optical Sensor Model

An optical sensor is a device designed to measure the intensity of optical energy (light) entering the sensor through an aperture. Imaging sensors consist of an array of light-sensitive detectors which create a two-dimensional light intensity measurement (i.e., image). In this section, the basic physical properties of an optical sensor are presented, and a model representing an optical sensor is given.

For the purposes of this discussion, the *world* is defined as a collection of all real objects. Some objects are sources of radiometric illumination or *radiance*. These light sources illuminate the world and interact with the other physical objects through various types of reflection. The amount of light along a certain direction is defined as the *irradiance* [13]. The physical irradiance pattern entering the aperture of the optical sensor is defined as the *scene* and is represented by a continuous array of nonnegative real numbers,  $o(x, y, t)$ , projected onto the image plane. For the purposes of this discussion, the irradiance sources are constrained to an arbitrary, piecewise continuous, Lambertian surface in three dimensions.

A digital optical imaging sensor consists of an aperture, lens, detector array, and sampling array. A simple imaging system model is shown in Figure 2. The lens focuses the scene on the detector array. The light pattern focused on the detector array is defined as the *image* and represented by,  $i(x, y, t)$ . In statistical terms, the *image* is the mean photon arrival rate, and is defined by a Poisson distribution [4]. The detector array converts the light energy into a voltage or a charge which is converted to a digital value by the sampling array. The sampling array is assumed to be a square grid, although other patterns can be designed (e.g., honeycomb) [8].

The lens is an analog low-pass filter in the spatial domain, with a cutoff frequency ( $f_c$ ) determined by the aperture ( $D$ ), wavelength of light source ( $\lambda$ ), and focal length



**Figure 2:** Digital imaging system. The imaging system transforms the scene into a digital image. The major components of the camera are the optics, light detector, amplifier, and analog to digital converter.

of the camera ( $f_0$ ) [4]:

$$f_c = \frac{D}{\lambda f_0} \left[ \frac{1}{m} \right] \quad (1)$$

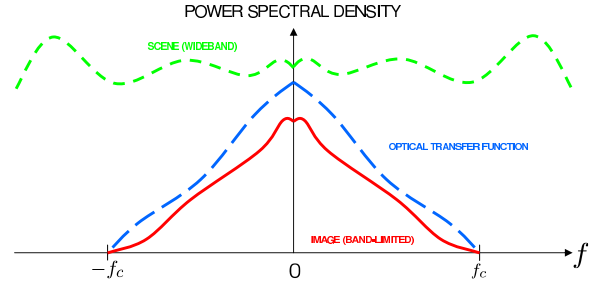
Thus, a scene consisting of a point source of light (delta function intensity) would appear slightly blurred (spread) on the image plane. Assuming spatial invariance, this blurring due to the lens is represented by the *point spread function*,  $h(\xi, \rho)$ , where  $\xi$  and  $\rho$  are the spatial differences in the  $x$  and  $y$  directions, respectively. The image in the spatial domain can now be expressed mathematically as the convolution of the scene and point spread function [6]

$$\mathbf{i}(x, y, t) = \int_{\xi \in \mathbf{X}} \int_{\rho \in \mathbf{Y}} \mathbf{o}(\xi, \rho, t) \mathbf{h}(x - \xi, y - \rho) d\rho d\xi \quad (2)$$

The image is physically continuous in space and time. This continuous function of three variables is then sampled and converted to an array of (digital) numbers. Concerning the sample process, the light energy in the image is integrated in each pixel over a temporal period defined as the *dwell time* ( $\Delta t$ ). The sampled image ( $\mathbf{i}_s(m, n, i)$ ) is obtained for integer pixel location ( $m, n$ ) and sample time,  $t_i$ , as

$$\mathbf{i}_s(m, n, t_i) = \int_{t_i - \Delta t/2}^{t_i + \Delta t/2} \int_{m - \Delta x/2}^{m + \Delta x/2} \int_{n - \Delta y/2}^{n + \Delta y/2} \mathbf{i}(x, y, t_i) dx dy dt \quad (3)$$

Analyzing the image sampling process from a frequency domain perspective provides insights into the rigorous use of images for navigation purposes. As previously mentioned, the camera optics act as an analog low-pass filter, characterized by the time-invariant point spread function,  $h(\xi, \rho)$ . The frequency domain representation of the point spread function,  $\mathbf{H}(f_x, f_y)$ , is called the *optical transfer function*. Applying the Fourier transform to the image equation (2), shows the multiplicative relationship for the



**Figure 3:** Effects of camera optics on image spatial frequency. The camera optics act as a low-pass filter with a cutoff frequency of  $f_c$ . The scene, which is wideband, appears as a band-limited image on the detector array.

spatial frequency domain representation of the image:

$$\mathbf{I}(f_x, f_y, t) = \mathbf{O}(f_x, f_y, t) \mathbf{H}(f_x, f_y) \quad (4)$$

In most conditions, the projected scene can be treated as a wideband function relative to the optical transfer function, (i.e.,  $f_{c_{scene}} \gg f_{c_{OTF}}$ ). This results in the following spatial frequency limitation of the projected image

$$\mathbf{I}(f_x, f_y, t) = 0, \quad \forall |f_x|, |f_y| > f_c \quad (5)$$

This relationship is expressed graphically in Figure 3.

The sampling operation can be represented by a zero-order hold (or sample-and-hold) process in the spatial domain and as a natural sampling process in the time domain (see [19]). The resulting frequency spectrum for the sampled image consists of *sinc*-weighted copies of the image frequency response, located at integer multiples of the spatial sampling frequency. An illustration is shown in Figure 4. Hence, to prevent spatial aliasing, that is, to avoid the “diffraction limit”, the spatial sampling rate must satisfy the spatial Nyquist condition in both dimensions, which is determined by the camera optics as:

$$f_x, f_y > 2f_c = 2 \frac{D}{\lambda f_0} \quad (6)$$

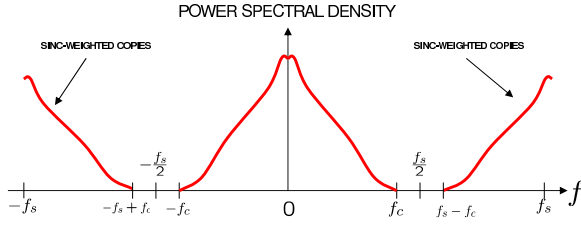
where  $f_x$  and  $f_y$  are the spatial sampling rates in the  $x$  and  $y$  directions, respectively. These are directly related to the physical pixel size as:

$$f_x = \frac{1}{\Delta x} \quad (7)$$

$$f_y = \frac{1}{\Delta y} \quad (8)$$

where  $\Delta x$  and  $\Delta y$  are the pixel sizes in the  $x$  and  $y$  directions.

Camera motion changes the projection of a stationary scene which, for a simple point illumination source, results in an apparent image “shift”. This image shift results in a modulation of the frequency content in the temporal frequency



**Figure 4: Spatial Sampling Illustration.** The spatial sampling process creates sinc-weighted spectral copies in the spatial frequency domain when square pixels are used. The sampling frequency,  $f_s$ , must be greater than twice the cut-off frequency,  $f_c$ , to eliminate spatial aliasing.

domain. More specifically, the velocity of a point source in the image plane,  $(\dot{s}_x^{proj}, \dot{s}_y^{proj})$ , results in the following Nyquist temporal sampling constraint

$$f_t > 2 \max \{ \dot{s}_x^{proj}, \dot{s}_y^{proj} \} f_c \quad (9)$$

Assuming square pixels which are sized according to the spatial Nyquist sampling (i.e.,  $f_x = f_y = 2f_c$ ) results in the following pixel size

$$\Delta_{pixel} = \frac{1}{2f_c} \quad (10)$$

Substituting Eqn. (10) into (9) results in the normalized temporal sampling constraint

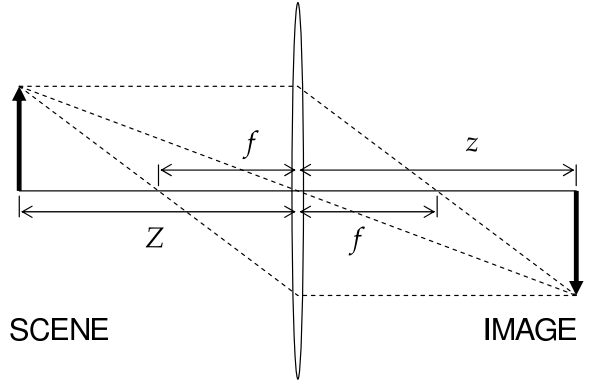
$$f_t > \frac{\max \{ \dot{s}_x^{proj}, \dot{s}_y^{proj} \}}{\Delta_{pixel}} \quad (11)$$

As a result, to minimize temporal aliasing, the Nyquist rate can be achieved by ensuring no feature moves more than one half of the minimum distance between intensity peaks in the image plane. Given an optical cutoff frequency of  $f_c$ , the temporal sampling interval,  $T_s$ , should be chosen such that the maximum image shift due to camera motion is less than  $\frac{1}{2f_c}$ . This implies a fundamental interrelationship between the minimum spatial and temporal sampling intervals, which is somewhat similar to the spatial-temporal discretization constraint found when solving the heat PDE, also known as the von Neumann condition.

In the next section, a mathematical model describing the relationship between point locations in the world and image will be derived. The resulting projection equations will be used to calculate appropriate temporal sampling intervals, based on scene geometry and camera motion.

### Egomotion Effects on Temporal Sampling

In the previous section, the effects of egomotion on the formation of the image are presented. In this section, the egomotion effects on temporal sampling are illustrated. Reducing the spatial dimensionality of the problem from two



**Figure 5: Thin lens camera model.** The thin lens model directs parallel light rays toward the focus, resulting in an image. Figure is not to scale.

to one is performed in order to illustrate the effects of egomotion on temporal sampling in a manner that is easier to visualize.

### Projection Theory

The camera optical properties define the relationship between the scene and the projected image. Recalling the simple camera model (Figure 2), the lens focuses the incoming irradiance pattern (i.e., scene) onto the image plane. For a theoretical thin lens, the projection is a function of the focal length of the lens and the distance from the lens, as shown in Figure 5. This relationship is expressed by the *fundamental equation of the thin lens* [13]:

$$\frac{1}{Z} + \frac{1}{z} = \frac{1}{f_0} \quad (12)$$

where  $Z$  is the distance from the object to the lens,  $z$  is the distance from the lens to the image plane, and  $f_0$  is the focal length.

As the aperture of the thin lens decreases to zero, the system can be modeled as a pinhole camera (see Figure 6). In this model, all incoming light must pass through the optical center and is projected on an image plane located at a distance  $f$  from the lens. The resulting image is an inverted projection of the scene.

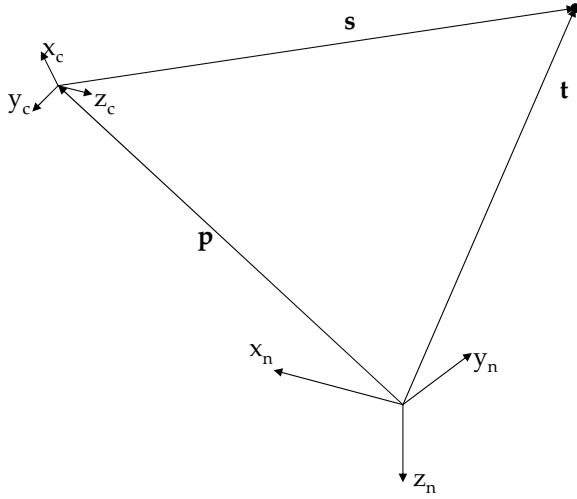
This model can be further simplified by placing a virtual image plane in front of the optical center, as shown in Figure 7. Given a point source at location  $\mathbf{s}^c$  the resulting location of the point source on the image plane, relative to the optical center of the camera, is given by

$$\mathbf{s}^{proj} = \left( \frac{f_0}{s_z^c} \right) \mathbf{s}^c = f_0 \underline{\mathbf{s}}^c \quad (13)$$

where  $s_z^c$  is the distance of the point source from the optical center of the camera in the  $z_c$  direction. The underline in-







**Figure 9:** Target to image transformation geometry. The relationship between the camera position, ( $\mathbf{p}$ ), and target location, ( $\mathbf{t}$ ), can be expressed in pixel coordinates using transformations based on the navigation state and camera parameters.

based on the navigation state. The geometry is shown in Figure 9. The line of sight vector,  $\mathbf{s}$ , is the vector difference between the target location  $\mathbf{t}$  and the camera position, which are both available in navigation frame coordinates ( $n$ ).

$$\mathbf{s}^n = \mathbf{t}^n - \mathbf{p}^n \quad (21)$$

The resultant vector can be transformed to the camera reference frame using the navigation-to-camera frame direction cosine matrix:

$$\mathbf{s}^c = \mathbf{C}_n^c \mathbf{s}^n \quad (22)$$

Finally, the pixel location is calculated using Eqn. (20).

### Apparent Pixel Motion Calculations

The previous development is extended to illustrate the apparent pixel motion of a point feature due to relative motion. The development begins by recalling the camera-to-pixel transformation shown in Eqs. (13-22).

$$\mathbf{s}^{pix} = \mathbf{T}_c^{pix} \underline{\mathbf{s}}^c = \mathbf{T}_c^{pix} \mathbf{s}^c / s_z^c \quad (23)$$

where the camera frame line of sight vector,  $\mathbf{s}^c$ , is given by

$$\mathbf{s}^c = \mathbf{C}_n^c [\mathbf{t}^n - \mathbf{p}^n] \quad (24)$$

The apparent pixel motion is derived by taking the derivative of  $\mathbf{s}^{pix}$  with respect to time:

$$\dot{\mathbf{s}}^{pix} = \mathbf{T}_c^{pix} \dot{\underline{\mathbf{s}}}^c \quad (25)$$

where

$$\dot{\underline{\mathbf{s}}}^c = \frac{s_z^c \dot{\mathbf{s}}^c - \mathbf{s}^c \dot{s}_z^c}{(s_z^c)^2} \quad (26)$$

The time-derivative of the camera frame line-of-sight vector is given by

$$\dot{\mathbf{s}}^c = \mathbf{C}_n^c \boldsymbol{\Omega}_{cn}^n [\mathbf{t}^n - \mathbf{p}^n] + \mathbf{C}_n^c [\dot{\mathbf{t}}^n - \dot{\mathbf{p}}^n] \quad (27)$$

where  $\boldsymbol{\Omega}_{cn}^n$  is the skew-symmetric form of the angular rate of the camera to the navigation frame, expressed in the navigation frame. The skew-symmetric form is defined in [23]. Expressing the rotations in the camera frame yields the following equivalent form:

$$\dot{\mathbf{s}}^c = -\boldsymbol{\Omega}_{nc}^c \mathbf{s}^c + \mathbf{C}_n^c [\dot{\mathbf{t}}^n - \dot{\mathbf{p}}^n] \quad (28)$$

Analysis of Eqn. (28) shows that the change in line-of-sight vector is a function of both the camera rotation and relative translational motion between the camera and landmark of interest.

In many cases, the landmark motion relative to the navigation frame is insignificant and can be neglected. Applying this assumption and coordinatizing the camera translational motion in the camera frame yields

$$\dot{\mathbf{s}}^c = -\boldsymbol{\Omega}_{nc}^c \mathbf{s}^c - \mathbf{v}^c \quad (29)$$

where  $\mathbf{v}^c$  is the velocity of the camera, relative to the navigation frame, coordinatized in the camera frame. Combining Eqs. (25), (26) and (29) results in the well-known optical flow equation [21]:

Substituting Eqn. (29) into Equations (26) and (25) results in the apparent pixel motion in the  $x$  and  $y$  directions:

$$\dot{u} = -\frac{f_0}{\Delta x} \left( -\omega_y - \frac{v_x}{s_z^c} + \frac{s_x^c s_y^c}{(s_z^c)^2} \omega_x - \left( \frac{s_x^c}{s_z^c} \right)^2 \omega_y + \frac{s_y^c}{s_z^c} \omega_z + \frac{s_x^c}{(s_z^c)^2} v_z \right) \quad (30)$$

$$\dot{v} = \frac{f_0}{\Delta y} \left( \omega_x - \frac{v_y}{s_z^c} - \frac{s_x^c s_y^c}{(s_z^c)^2} \omega_y + \left( \frac{s_y^c}{s_z^c} \right)^2 \omega_x - \frac{s_x^c}{s_z^c} \omega_z + \frac{s_y^c}{(s_z^c)^2} v_z \right) \quad (31)$$

which is expressed using the scalar components of the rotation, velocity, and line-of-sight vectors referenced in Equation (29).

The temporal sampling constraint proposed in the previous section indicates that it is desirable to sample such that the apparent pixel motion is limited to no more than one pixel of change per image in both the  $x$  and  $y$  spatial dimensions, provided the image is sampled at spatial Nyquist frequency. Given a sample interval,  $T_s$ , the maximum pixel motion component,  $K_{max}$  can be approximated by

$$K_{max} = \max \{ |\dot{u}| T_s, |\dot{v}| T_s \} \leq 1 \quad (32)$$

In the next section, the derived apparent pixel motion is analyzed for a representative scenario which illustrates the difficulty in achieving samples from traditional imaging systems which do not violate the temporal sampling constraints presented above.

## ILLUSTRATIVE CASE STUDY

In this section, the apparent pixel motion is calculated for a selection of representative imaging scenarios. As previously developed, the generalized sampling characteristics of a given imaging sensor is a function of a number of parameters. In this scenario, we will assume that the camera intrinsic parameters (i.e.,  $\Delta x$ ,  $\Delta y$ ,  $f_0$ ,  $D$ , and  $\lambda$ ) are fixed in such a way to guarantee proper spatial sampling. For this case, we are interested in the resulting temporal sampling rate ( $f_t$ ) which is consistent with the temporal sampling constraints derived in the previous section. The camera intrinsic parameters are chosen to be representative of currently available machine-vision cameras. These parameters are shown in Table 1.

**Table 1: Camera Intrinsic Parameters.** The camera intrinsic parameters are chosen to be representative of currently available machine vision cameras and are chosen to eliminate spatial aliasing.

Description	Parameter	Value	(Units)
Wavelength	$\lambda$	550	$\mu m$
Focal length	$f_0$	6	$mm$
Lens Aperture	$D$	6/16	$mm$
Vertical Image Size	$M$	1024	$pixels$
Vertical Pixel Size	$\Delta x$	4.4	$\mu m$
Horizontal Image Size	$N$	1280	$pixels$
Horizontal Pixel Size	$\Delta y$	4.4	$\mu m$

The first case study is a simple  $5 \frac{deg}{sec}$  horizontal pan, with no translational motion. The resulting motion parameters for this condition are as follows:

$$\mathbf{v}^c = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \left( \frac{m}{s} \right) \quad (33)$$

$$\boldsymbol{\omega}_{nc}^c = \begin{bmatrix} 5 \frac{\pi}{180} \\ 0 \\ 0 \end{bmatrix} \left( \frac{rad}{sec} \right) \quad (34)$$

Substituting these motion parameters and the intrinsic camera parameters into Eqs. (30) and (31) yields

$$K_{max} = \frac{6mm}{4.4\mu m} T_s \max \left\{ \left| \frac{s_x^c s_y^c}{(s_z^c)^2} \right|, \left| 1 + \left( \frac{s_y^c}{s_z^c} \right)^2 \right| \right\} \frac{5\pi}{180} \quad (35)$$

As evident in Eqn. (35), the pixel motion is primarily a function of the camera motion with second-order effects related to the position of the point source within the image. The worst-case condition occurs at the extreme extents of the image. Substituting these conditions into (35) resolves the maximization ambiguity

$$\begin{aligned} K_{max} &= \frac{6mm}{4.4\mu m} T_s \max \{0.1762, 1.2203\} \frac{5\pi}{180} \\ &= 145.2 T_s \end{aligned} \quad (36)$$

Applying the temporal sampling constraint and solving for  $T_s$  yields:

$$T_s \leq \frac{1}{145.2} \quad (sec) \quad (38)$$

which results in a minimum frame rate of 145.2 Hz and, consequently, a maximum exposure time of 6.9 ms.

In the next example, the effects of translational motion are investigated. Here, the camera is moving at 300 meters per second with a fixed orientation. The distance to the terrain is 10,000 meters, which represents a high-altitude cruise profile for an aircraft. The resulting motion parameters for this condition are as follows:

$$\mathbf{v}^c = \begin{bmatrix} 300 \\ 0 \\ 0 \end{bmatrix} \left( \frac{m}{s} \right) \quad (39)$$

$$\boldsymbol{\omega}_{nc}^c = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \left( \frac{rad}{sec} \right) \quad (40)$$

Substituting these motion parameters and the intrinsic camera parameters into Eqs. (30) and (31) yields

$$K_{max} = \frac{6mm}{4.4\mu m} T_s \max \left\{ \frac{300 \frac{m}{s}}{10000m}, 0 \right\} \quad (41)$$

$$K_{max} = 40.9 T_s \quad (42)$$

Applying the temporal sampling constraint and solving for  $T_s$  yields:

$$T_s \leq \frac{1}{40.9} \quad (sec) \quad (43)$$

which results in a minimum frame rate of 40.9 Hz and maximum exposure time of 24.4 ms.

The final example represents the conditions expected during a low-level, high speed dash profile. As in the previous example, the camera is moving at 300 meters per second with a relatively fixed orientation. However in this case, the distance to the terrain is reduced to 300 meters. The resulting motion parameters for this condition are as follows:

$$\mathbf{v}^c = \begin{bmatrix} 300 \\ 0 \\ 0 \end{bmatrix} \left( \frac{m}{s} \right) \quad (44)$$

$$\boldsymbol{\omega}_{nc}^c = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \left( \frac{rad}{sec} \right) \quad (45)$$

Substituting these motion parameters and the intrinsic camera parameters into Eqs. (30) and (31) yields

$$K_{max} = \frac{6mm}{4.4\mu m} T_s \max \left\{ \frac{300 \frac{m}{s}}{300m}, 0 \right\} \quad (46)$$

$$K_{max} = 1363.6 T_s \quad (47)$$

Applying the temporal sampling constraint and solving for  $T_s$  yields:

$$T_s \leq \frac{1}{1363.6} \quad (sec) \quad (48)$$

which results in a minimum frame rate of 1363.6 Hz and maximum exposure time of 733  $\mu s$ .

These case studies illustrate the frame rates required to sample at the Nyquist frequency. In general, the desired frame rates are not readily attainable using common hardware and lighting conditions. In many current correspondence search schemes, (e.g. [11], [9], [12]), the Nyquist sampling frequency for point sources is simply ignored and the search scheme seeks so-called “strong” features which are consistent between frames and geometrically consistent within a collection of other features (e.g., RANSAC). It is our assertion that these feature extraction and correspondence techniques are effectively applying low-pass, anti-aliasing filters which eliminate the higher frequency components which are corrupted by temporal aliasing.

As mentioned previously, there is a strong coupling between changes in camera pose and the apparent pixel motion. In the next section, measurements from an inertial sensor are used to mitigate the effects of temporal aliasing.

## INCORPORATION OF INERTIAL SENSOR MEASUREMENTS

As shown in the previous section, non-aliased temporal sampling can require relatively high frame rates, even for relatively simple imaging scenarios. High frame rates can present a number of challenges for a given imager, including high communication bandwidth requirements and short exposure times, requiring more sensitive (and expensive) sensors. We propose to exploit the information provided by inertial sensors in order to reduce the image sampling rates required to deliver anti-aliased measurements. The development of the aided sampling theory is presented as follows.

Inertial sensors can provide three-dimensional measurements of both angular rate and specific force (i.e., the sum of acceleration with respect to inertial and gravity) [22]. When combined with a kinematic model, this information can be exploited to produce an estimate of trajectory. For the purposes of this illustration, the error dynamics can be sufficiently modeled using the following method.

When target motion was assumed to be effectively station-

ary, the apparent pixel motion (Eqns. 30 and 31) was a function of the camera rotation rate and velocity with respect to the navigation frame and the relative location of the landmark. Strapdown inertial sensors measure both the angular rotation increment,  $\Delta \theta_{ic}^c$ , and specific force increment,  $\Delta \mathbf{v}^c$ , with respect to the inertial reference frame. When combined with knowledge of the gravity vector, kinematic equations can be used to estimate the position, velocity, and attitude of the sensor. The inertial measurement errors, initial navigation state uncertainty, and errors in the gravity model all contribute to the inevitable, unstable error growth experienced by all unaided strapdown inertial navigation systems. A thorough development of these properties can be found in [22].

While all inertial navigation systems experience unstable error growth over time, the relatively short durations between images allow us to model the errors between successive images using a simpler model. The first approximation assumes that the navigation reference frame is effectively an inertial reference frame over the short term. The second approximation assumes a general knowledge of the navigation state (e.g., the system is reasonably aligned) such that any errors in the navigation state itself do not dominate the pixel motion prediction between frames.

Thus, the simplified inertial sensor model represents the measurement as the sum of the true value plus an error and is given as

$$\tilde{\omega}_{nc}^c = \omega_{nc}^c + \delta \omega_{nc}^c \quad (49)$$

$$\tilde{\mathbf{v}}^c = \mathbf{v}^c + \delta \mathbf{v}^c \quad (50)$$

where  $\omega_{nc}^c$  is the true angular rotation rate and  $\mathbf{v}^c$  is the true velocity. The tilde represents the corrupted measurement as received from the inertial sensor. The inertial measurement errors,  $\delta \omega_{nc}^c$  and  $\delta \mathbf{v}^c$  can be represented as random vectors with the following statistics over the interval  $T_s$ :

$$\mathbf{E} [\delta \omega_{nc}^c] = \mathbf{0}_{3 \times 3} \quad (51)$$

$$\mathbf{E} [\delta \omega_{nc}^c \delta \omega_{nc}^{cT}] = q_w \quad (52)$$

$$\mathbf{E} [\delta \mathbf{v}^c] = \mathbf{0}_{3 \times 3} \quad (53)$$

$$\mathbf{E} [\delta \mathbf{v}^c \delta \mathbf{v}^{cT}] = (\sigma_{v_0}^2 + q_a T_s) \mathbf{I}_{3 \times 3} \quad (54)$$

The gyroscopic and accelerometer error sources are assumed to be collectively independent. Substituting the velocity and angle increment measurements from the inertial sensor algorithm into the pixel motion equations from Eqns. (30) and (31) and integrating the error terms results in the residual pixel motion error rate due to inertial mea-

surement errors:

$$\delta \dot{u} = -\frac{f_0}{\Delta x} \left( -\delta \omega_{nc_y}^c - \frac{\delta v_x^c}{s_z^c} + \frac{s_x^c s_y^c}{(s_z^c)^2} \delta \omega_{nc_x}^c \right. \\ \left. - \left( \frac{s_x^c}{s_z^c} \right)^2 \delta \omega_{nc_y}^c + \frac{s_y^c}{s_z^c} \delta \omega_{nc_z}^c + \frac{s_x^c}{(s_z^c)^2} \delta v_z^c \right) \quad (55)$$

$$\delta \dot{v} = \frac{f_0}{\Delta y} \left( \delta \omega_{nc_x}^c - \frac{\delta v_y^c}{s_z^c} - \frac{s_x^c s_y^c}{(s_z^c)^2} \delta \omega_{nc_y}^c \right. \\ \left. + \left( \frac{s_y^c}{s_z^c} \right)^2 \delta \omega_{nc_x}^c - \frac{s_x^c}{s_z^c} \delta \omega_{nc_z}^c + \frac{s_y^c}{(s_z^c)^2} \delta v_z^c \right) \quad (56)$$

where  $\delta \dot{u}$  and  $\delta \dot{v}$  are the random pixel location errors rates in the  $x$  and  $y$  directions, respectively. The standard deviation of the residual pixel errors is given by calculating the variance of pixel errors after integrating over an interval of  $T_s$ , yielding:

$$\sigma_u = \frac{f_0}{\Delta x} \left[ T_s \left( q_w + \frac{\sigma_{v_0}^2 + q_a T_s}{s_z^c} + \frac{s_x^c s_y^c}{(s_z^c)^2} q_w \right. \right. \\ \left. \left. + \left( \frac{s_x^c}{s_z^c} \right)^2 q_w + \frac{s_y^c}{s_z^c} q_w + \frac{s_x^c (\sigma_{v_0}^2 + q_a T_s)}{(s_z^c)^2} \right) \right]^{1/2} \quad (57)$$

$$\sigma_v = \frac{f_0}{\Delta y} \left[ T_s \left( q_w + \frac{\sigma_{v_0}^2 + q_a T_s}{s_z^c} + \frac{s_x^c s_y^c}{(s_z^c)^2} q_w \right. \right. \\ \left. \left. + \left( \frac{s_y^c}{s_z^c} \right)^2 q_w + \frac{s_x^c}{s_z^c} q_w + \frac{s_y^c (\sigma_{v_0}^2 + q_a T_s)}{(s_z^c)^2} \right) \right]^{1/2} \quad (58)$$

The temporal sampling constraint can be applied in a similar manner as before, however in this case the constraint is applied to the standard deviation of *residual error* of pixel motion versus the total pixel motion considered in the unaided case.

$$\sigma_{K_{max}} = \max \{ \sigma_u, \sigma_v \} \quad (59)$$

Enforcing the temporal sampling constraint on the residual random pixel motion requires selecting a confidence interval such that the residual pixel motion is constrained to less than one pixel uncertainty. This can be accomplished by evaluating the resulting probability distribution function of the residual pixel errors.

The preceding development is illustrated using a simple example. In this example, a consumer-grade inertial sensor is available with the following random walk parameters:

$$q_w = 4.2 \times 10^{-7} \quad \frac{rad^2}{s} \quad (60)$$

$$q_a = 1.9 \times 10^{-5} \quad \frac{\left( \frac{m}{s} \right)^2}{s} \quad (61)$$

As a further simplification, the pan components are isolated by assuming relatively distant targets (e.g.,  $s_z^c \rightarrow \infty$ ). This

results in the following pixel uncertainties

$$\sigma_u = \frac{f_0}{\Delta x} [T_s q_w]^{1/2} \quad (62)$$

$$\sigma_v = \frac{f_0}{\Delta y} [T_s q_w]^{1/2} \quad (63)$$

Applying a 3- $\sigma$  bound to the prediction errors results in the following temporal sampling constraint

$$3\sigma_{K_{max}} = 3 \max \left\{ \frac{f_0}{\Delta x} [T_s q_w]^{1/2}, \right. \\ \left. \frac{f_0}{\Delta y} [T_s q_w]^{1/2} \right\} \leq 1 \quad (64)$$

$$3\sigma_{K_{max}} = 3f_0 [T_s q_w]^{1/2} \\ \max \left\{ \frac{1}{\Delta x}, \frac{1}{\Delta y} \right\} \leq 1 \quad (65)$$

Solving for the sampling interval yields

$$T_s \leq \frac{1}{9f_0^2 q_w} \min \{ \Delta x^2, \Delta y^2 \} \quad (66)$$

Substituting the previously presented camera and inertial parameters yields:

$$T_s \leq \frac{1}{7.03} \quad (sec) \quad (67)$$

which results in a minimum frame rate of 7.03 Hz and maximum exposure time of 142.3 ms.

This illustration shows the benefits possible when utilizing inertial measurements to reduce temporal aliasing. This result implies that as long as the rotational motion is within the bandwidth of the inertial sensor, sampling at  $\geq 7.03$  Hertz will give acceptable anti-aliased results. Thus, when incorporating inertial measurements for feature anti-aliasing, the sampling rate is independent of camera motion.

## CONCLUSIONS AND FUTURE WORK

In this paper, the concepts relating spatial and temporal image sampling are explored from first principles with focus on the consequences for the correspondence search and feature tracking problem. The sampling theory is developed and shown to yield a natural interrelationship between acceptable spatial and temporal sampling frequencies. The relationships between apparent feature motion and temporal sampling requirements are shown to require very high (possibly unattainable) temporal sampling rates in order to guarantee un-aliased sampling. We believe this is an underlying cause which forces designers to exploit complicated feature tracking algorithms, which, in essence, can be viewed as sophisticated anti-aliasing filters. A case in

point is the operation of the optical mouse. It uses a correlation algorithm and assumes a linear ( $x$ - $y$ ) planar motion. This, in turn, simplifies the computations and allows for the use of a very high sampling rate: evidently, the aliasing/ambiguity issue is well appreciated because the sampling rate used is 1.8 kHz [2].

Once the problem is posed from this perspective, the incorporation of inertial sensors is a natural choice. Inertial sensors are shown to have the capability to statistically constrain the apparent motion effects, which can result in a significant reduction in required temporal sampling rates while alleviating the burden of feature correspondence search. In essence, inertial sensors are proposed to provide us with a direct method for reducing or eliminating temporal aliasing, allowing for the use of sophisticated and efficient/robust correspondence search algorithms and operation under lower-lighting conditions.

Indeed, the use of inertial measurements for aiding the feature correspondence search task is akin to the use of inertial measurements in ultra-tightly coupled GPS and INS where the inertial information is used to steer the phase-locked loops in a feed-forward mechanization. This facilitates precise code-tracking under dynamic conditions – a powerful combination of precision and robustness which is the hallmark of properly fused synergistic sensors [15].

There are a number of issues which require further work and development. First, applying statistical constraints from inertial sensors requires some knowledge of the scene to properly account for translational motion. We propose to address this issue by incorporating statistical knowledge of the terrain (either *a priori* or *in situ*) which could be applied dynamically to either control temporal sampling rate or to exclude features for which aliasing is predicted.

Secondly, this development does not exploit any geometric constraints regarding the scene itself. In certain cases, (e.g., an aircraft imaging a relatively flat scene) the temporal sampling rate can be reduced below the worst-case threshold presented in this paper.

Ultimately, we believe this theory demonstrates the complementary nature of imaging and inertial sensors. As such, properly incorporating inertial sensors can be a major advantage in developing robust image tracking applications within reasonable imaging and image processing constraints.

## DISCLAIMER

The views expressed in this article are those of the author and do not reflect the official policy or position of the United States Air Force, Department of Defense, or the U.S. Government.

## References

- [1] Henele Adams, Sanjiv Singh, and Dennis Strelow. An empirical comparison of methods for image-based motion estimation. In *Proceedings of the 2002 IEEE/RSJ Intl. Conference on Intelligent Robots and Systems*, volume 1, pages 123–128, September 2002.
- [2] Agilent Technologies Inc. Agilent adns-2610 optical mouse sensor. Datasheet, October 2004.
- [3] Bir Bhanu, B. Roberts, and J. Ming. Inertial navigation sensor integrated motion analysis for obstacle detection. In *Proceedings IEEE International Conference on Robotics and Automation*, pages 954–959, May 1990.
- [4] Joseph W. Goodman. *Introduction to Fourier Optics*. McGraw Hill, Boston, Massachusetts, 1996.
- [5] Espen Hagen and Eilert Heyerdahl. Navigation by optical flow. In *Proceeding of the 11th IAPR International Conference on Pattern Recognition*, volume 1, pages 700–703, 1992.
- [6] Eugene Hecht. *Optics*. Addison Wesley, San Francisco, California, 2002.
- [7] Stefan Hrabar and Gaurav S. Sukhatme. A comparison of two camera configurations for optic-flow based navigation of a UAV through urban canyons. In *Proceedings of the 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems*, volume 3, pages 2673–2680, September 2004.
- [8] Anil K. Jain. *Fundamentals of Digital Image Processing*. Prentice Hall, Inc., Upper Saddle River, New Jersey 07458, 1989.
- [9] Yan Ke and Rahul Sukthankar. Pca-sift: A more distinctive representation for local image descriptors. In *2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'04)*, volume 2, pages 506–513, 2004.
- [10] David G. Lowe. Object recognition from local scale-invariant features. In *Proc. of the International Conference on Computer Vision*, volume 2, pages 1150–1157, September 1999. Corfu, Greece.
- [11] David G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.
- [12] B. D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. *Proc. DARPA Image Understanding Workshop*, pages 121–130, 1981.

- [13] Yi Ma, Stefano Soatto, Jana Kosecka, and S. Shankar Sastry. *An Invitation to 3-D Vision*. Springer-Verlag, Inc., New York, New York, 2004.
- [14] Clark F. Olson, Larry H. Matthies, Marcel Schoppers, and Mark W. Maimone. Robust stereo ego-motion for long distance navigation. In *Proceedings of the IEEE Conference on Advanced Robotics*, volume 2, pages 453–458, June 2000.
- [15] Mark E. Oxley and Steven N. Thorsen. Fusion or integration: What’s the difference? In *7th International Conference on Information Fusion (FUSION 2004)*, pages 429–434, June 2004.
- [16] Meir Pachter and Alec Porter. Bearings-only measurements for INS aiding: The three-dimensional case. In *Proceedings of the 2003 AIAA Guidance, Navigation and Control Conference*, 2003. AIAA paper number 2003-5354.
- [17] Meir Pachter and Alec Porter. INS aiding by tracking an unknown ground object - theory. In *Proceedings of the American Control Conference*, volume 2, pages 1260–1265, 2003.
- [18] John F. Raquet and Michael Giebner. Navigation using optical measurements of objects at unknown locations. In *Proceedings of the 59th Annual Meeting of the Institute of Navigation*, pages 282–290, June 2003.
- [19] Bernard Sklar. *Digital Communications: Fundamentals and Applications*. Prentice-Hall PTR, Upper Saddle River, NJ, 2000.
- [20] Dennis W. Strelow. *Motion Estimation from Image and Inertial Measurements*. PhD thesis, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213, November 2004.
- [21] A. Murat Tekalp. *Digital Video Processing*. Prentice Hall PTR, Upper Saddle River, New Jersey 07458, 1995.
- [22] D.H. Titterton and J.L. Weston. *Strapdown Inertial Navigation Technology*. Peter Peregrinus Ltd., Lavenham, United Kingdom, 1997.
- [23] Michael J. Veth and John F. Raquet. Alignment and calibration of optical and inertial sensors using stellar observations. In *Proceedings of ION GNSS 2005*, pages 2494–2503, September 2005.
- [24] Michael J. Veth and John F. Raquet. Two-dimensional stochastic projections for tight integration of optical and inertial sensors for navigation. In *National Technical Meeting Proceedings of the Institute of Navigation*, pages 587–596, 2006.
- [25] Michael J. Veth, John F. Raquet, and Meir Pachter. Stochastic constraints for efficient image correspondence search. *IEEE Transactions on Aerospace Electronic Systems*, 42(3):973–982, July 2006.